

## **Demand Forecasting and Management for Container Volume in Northeast Marmara**

**Yusuf BAYKAN<sup>1</sup>**  
**Murat YORULMAZ<sup>2</sup>**  
**Mert SUSOY<sup>3</sup>**  
**Ayşe TAŞ<sup>4</sup>**

Received: 25.01.2026, Accepted: 03.05.2026  
DOI Number: 10.5281/zenodo.20563222

### **Abstract**

Demand forecasting plays a critical role in sustaining port operations, ensuring traffic order, preventing security risks, and enabling the efficient allocation of resources in ports, which constitute the backbone of maritime transportation within the global logistics system. This study aims to forecast future container demand in the Northeastern Marmara Region by examining the effects of key macroeconomic indicators, namely gross domestic product (GDP), population, and foreign trade balance, and to provide strategic insights for port capacity planning and sectoral development. To achieve this objective, data from 2012 to 2021 was used, and panel data regression analysis was employed to identify relationships between container demand and the selected variables. At the same time, time-series forecasting techniques are applied to predict future cargo and container volumes. In this context, ARIMA, Exponential Smoothing (ETS), TBATS, Extreme Learning Machine (ELM), and Multilayer Perceptron (MLP) neural network models are comparatively analyzed using the R statistical software. The empirical findings reveal that GDP has a statistically significant and positive effect on both total cargo and container volumes. In contrast, population and the foreign trade balance have a negative or statistically insignificant effect. Forecasting results indicate a substantial increase in container handling volumes at the Northeastern Marmara ports over the next five-year period, with the TBATS model providing the most accurate predictions among the methods applied. The study contributes to the literature by integrating panel data analysis with multiple forecasting techniques. It offers practical implications for port authorities and policymakers in developing capacity expansion strategies and improving demand-oriented port management.

---

<sup>1</sup>MSc, Kocaeli University, Türkiye, yusufbaykan1988@gmail.com <https://orcid.org/0000-0002-7901-2500>

<sup>2</sup> Assoc. Prof. PhD., Kocaeli University, Türkiye, murat.yorulmaz@kocaeli.edu.tr <https://orcid.org/0000-0002-5736-9146>

<sup>3</sup>MSc, Kocaeli University, Türkiye, susoy94@gmail.co, <https://orcid.org/0009-0000-3383-8545>

<sup>4</sup>Asist. Prof. PhD., Kocaeli University, Kocaeli, Türkiye, ayse.tas@kocaeli.edu.tr, <https://orcid.org/0000-0002-2457-7723>

**Key words:** Maritime Transportation, Ports, Container Demand, Panel Data Analysis, Demand Forecasting.

**JEL Code:** C23, C53, R41, L91.

## **1. Introduction**

Consideration of increasing world population and globalization, it is reasonable to claim that countries only survive by producing for exportation. In today's world almost 90 percent of the goods subject to international trade are transported by sea. This ratio simply implies that maritime trade constitutes a large part of world trade. Ports represent the only gateway of maritime transport to international trade.

Demand forecasting at ports, which are a sine qua non of international trade, provides advantages in allocating resources, preventing waste of resources, establishing sales policies, pricing, reducing business risk, and managing inventory. Thus, businesses can gain insights and predict their future.

The concept of container transportation was first introduced by Malcolm P. McLean, and the first ship named the Maxton, converted from a tanker ship, was used to carry 60 containers. Containers have been used effectively for international maritime transportation for more than half a century (Solmaz & Saygılı, 2008) and their use has been on the rise as they are based on the concept of a unit cargo.

According to the 2022 TÜRKLİM (Port Operators Association of Türkiye) Industry Report, the container handled throughout the world was 542 million tons, and Türkiye handled 4.5% of the world's containers, a total of 542 million tons, through ports located within its borders. The seven ports located in the northeastern Marmara hinterland take over 86 million tons of the country's total handling. More specifically, regarding container transportation in Türkiye, 27 ports in the country handle containerized cargo (TÜRKLİM, 2022). Among these ports, Haydarpaşa, Beldeport, DP WORLD, Evyapport, LİMAŞ, Safiport, and Yılport are the ports located in the northeastern Marmara hinterland containing the container terminal. If we look at these ports' capacities, Haydarpaşa Port serves with an annual handling capacity of 655 thousand TEU; Beldeport Port has 550 thousand TEU; DP World Port has 1 million 300 thousand TEU; Evyapport Port, 855 thousand TEU; LİMAŞ Port, 250 thousand TEU; Safiport Derince Port, 1 million 500 thousand TEU; Yılport Container Port, 1 million TEU (TÜRKLİM, 2022). Haydarpaşa, Beldeport, DP WORLD, Evyapport, LİMAŞ, Safiport, and Yılport container ports, which are in the Northeastern Marmara hinterland that is the subject of the study, are pivotal for the national economy and maritime transportation in terms of the container capacity of Türkiye.

This study benefits from the R software for the analysis part, since it is one of the most extensive statistical analysis software. Today, R software is one of the most popular programming languages and is favored in the private sector besides academic studies. For conceptual clarity, this study distinguishes between key terms

used throughout the analysis. “Container demand” refers to the forecasted levels of container handling, whereas “container volume” represents the observed historical throughput data used in the modeling process. In addition, “cargo demand” is used as a broader term encompassing both containerized and non-containerized cargo flows.

## 2. Literature Review

If we look at studies regarding the estimation of cargo demands in ports in the literature, TÜRKLİM studies the data obtained and offers regional forecasts for the coming years. According to the latest TÜRKLİM report that draws on the regression technique, an 11% growth is projected in the northeastern Marmara region in 2021 (TÜRKLİM, 2022). Akar and Esmer (2015) conducted multiple linear regression analyses to perform the demand forecasting for cargo and determine whether the ports in Türkiye are qualified enough to handle this cargo. Their study concluded that cargo handling in 2023 will be equal to 12.7 million TEU (Akar & Esmer, 2015).

Ubaid et al. (2021) dealt with the demand forecasting for the Asian and Oceania trade routes and predicted the change in imports through certain techniques including a time series. This study considers import and export figures together to prevent calculation errors that may arise due to the dependent variables in the analysis. Five different methods, namely, ARIMA, ETS-Exponential Deceleration Model, ELM Artificial Neural Networks, MLP Artificial Neural Networks, and TBATS model, were used. The reason for using several methods is to identify the most suitable model for forecasting, and it was found that each region and group of data required a different method for analysis (Ubaid, Hussain, & Saqib, 2021). According to the Final Report of the Coastal Structures Master Plan by the Ministry of Transportation, GDP (Gross Domestic Product), wholesale or consumer price index, personal consumption expenditures, capacity utilization rates, transport volumes of raw materials, semi-finished and finished products, import and export volumes, regional consumption of basic products are used as data for demand forecasting. Demand forecasting for cargo volume for ports usually benefits from multiple, linear, and logarithmic regression models, and the amount of cargo, population, and GDP data are used as independent variables.

Kara (2011) carried out demand forecasting for the future cargo traffic of İzmir (Alsancak) Port by 2016 using the time series method based on handling data on containers and general cargo at İzmir (Alsancak) Port. The researcher further drew on an autoregressive integrated moving average (ARIMA) method developed by Box and Jenkins in the time series analysis, which performed 60-month data forecasts based on 12-month forward and backward periods. It was ascertained that the İzmir Port experienced a lower increase compared to the increase in the total amount of containers handled across Türkiye and that the main reason for this and also for the decrease in the total amount of cargo handled is not the global crisis, but the low berth depth, lack of modernization, and insufficient storage. Mainardi and Santos (2016) estimated the cargo handling volumes at Portuguese ports using

linear regression through industrial and economic indicators from 2001 to 2014. Reliable forecasts of port output are extremely significant for ports and given the high investment and long time required to improve port infrastructure and superstructure, a good balance between port development and expected efficiency is a prerequisite. An oversized port will lead to delays in the cargo loading and unloading process and discourage shipowners from coming back to the port. Therefore, adequate port development entails reliable forecasts of port output. It is reported that the value of containerized cargo handling will increase from 2.5 million TEU to about 4.6 million TEU by 2025, and this will result in a need to increase capacity.

Using the multiple regression method for Kocaeli ports, Doğusel (2021) performed demand forecasting for the future total cargo tonnage and the total container amount in the region in the coming period. The data on total cargo and handled containers in Türkiye from 2009 to 2020 are expressed as dependent variables, while GDP, import, export, and population figures are treated as independent variables. Drawing on these variables, the researcher concluded that the capacity of Kocaeli ports will not have problems in meeting demand in regard to the total cargo, but the container capacity of 3.7 million TEU will fail to satisfy demand by 2033, so there is a need to boost container capacity in Kocaeli.

Kılınç et al. (2022) investigated the amount of container and cargo handling at the ports in Türkiye through the Artificial Neural Networks (ANNs) Nonlinear Autoregressive Network with External Input (NARX) Model. Their study benefited from two dependent variables, container handling and cargo handling volumes, as well as various independent variables including the Basket of Currencies Rate (USD-EURO), Gross Domestic Product (GDP), Consumer Confidence Index, Brent Oil, Export, Import and Industrial Production Index for forecasting. The data set consisted of the monthly data for the period between January 2004 and October 2020. The researchers also calculated the 18-month forecast values of container and cargo handling volumes for July, 2020-December 2021. Based on these predictive values and the announced values, a comparison between forecast and actual values was made for the first 4 months. Further, MAPE, MAE, RMSE, and MAD performance measures were calculated. The NARX neural network model provides very similar results to the real results, as it allows working with multiple variables and predicting nonlinear relationships of variables that have lost stability with high precision (Kılınç, Karaatlı, & Ömürbek, 2022).

### **3. Methodology**

The selection of forecasting models in this study is based on their ability to capture different characteristics of time series data. Traditional models such as ARIMA and ETS are widely used for linear trend and seasonality analysis, whereas TBATS is particularly suitable for handling complex seasonal patterns. In addition, artificial neural network models, including ELM and MLP, are employed to capture nonlinear relationships among variables and enhance predictive performance. The combined use of these models enables a comprehensive comparison and enhances

the robustness of the forecasting results. This paper is based on the first author's MSc thesis completed at Kocaeli University.

### 3.1. Regression Model for Panel Data Analysis

This study investigates the total cargo and container volumes of the ports located in the Northeastern Marmara hinterland using the population, GDP, import and export data of the provinces of the Marmara Region obtained from the TÜİK (Turkish Statistical Institute) and TÜRKİİM. Aktas (2010) revealed that imports and exports are related to each other, and any change in one of them affects the other (Aktas, 2010). Given that previous studies show that import and export data are interdependent variables, the foreign trade deficit obtained from these data was used in creating the model. The analysis method applied was the panel regression model. Equation 1 shows the formula obtained for the panel regression model, and information on the variables is given item by item. The model that this study uses compiled and analyzed the 2012-2021 data.

$$Y_{it} = \alpha_{it} + \beta_{it}X_{it} + \mu_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (1)$$

In the equation used,

Panel data analysis is more advantageous than other analyses applied in literature. The main reason for this is that it analyzes time series by integrating cross-sectional analysis. The advantages of panel data are as follows (Baltagi, 2015):

Cross-sectional data and time series data are integrated. This allows access to more observations and data. An increased number of observations brings about an increased degree of freedom. Thus, the connection problems within variables are reduced, and the efficiency and reliability of the prediction become greater. Considering the data to be used for this study, panel data appears as a more successful method of expressing the impact of more variables. As for econometric analyses, panel data analysis reveals more impact. This analysis, which is also compatible with a complex behavioral model, reduces the deviation values caused by cross-section to a minimum while using more data.

This study utilizes panel data analysis, which evaluates the time series together with the cross-sectional data, in investigating the ports of the Northeastern Marmara hinterland, which enjoys significant and abundant data on cargo in Türkiye.

Panel data models are divided into classes based on the criteria they contain. A panel is called a "balanced panel" if each unit contained in the data has been observed in all periods, and an "unbalanced panel" if not. Both models differ from each other. Panel data models are also divided into static and dynamic panel data based on the data they contain.

In panel data analysis, the Classical Model is a model in which both constant and slope coefficients are constant by unit and time (Yerdelen, 2017). In this

formula  $\beta_0$ , 0 is a fixed parameter and does not change depending on the cross-section or time series. The slope parameter of each unit, with the number of independent variables being K, is independent of the cross-section and time series in this unit.  $\beta_k$  In the broader sense, the formulation of the fixed effects model is similar to the basic formulation of panel data, and it is as shown in the following figure.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \dots + \gamma \cap D \cap I + \mu_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (2)$$

In the random effects model, it is not shown in a fixed parameter due to the lack of a fixed unit effect. This is expressed in the formula as the unit effect margin of error, and the error term is shown as follows.

$$v_{it} = \varepsilon_{it} + \mu_{it} \quad (3)$$

The integration of cross-sectional data and time series data in the analysis allows this study to better assess the impact of the volume of cargo handled at the Northeastern Marmara Ports by population, GDP, and foreign trade deficit. Necessary assumptions such as stationarity, model detection, variance variability, autocorrelation, and inter-unit correlations must be examined to obtain the correct results in the step of identifying the coefficients of the variables and determining their effect on the equation.

The formula of the random effects model obtained through the application of differences is expressed as follows (Yaffee, 2003).

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + u_{it} + \mu_{it} \quad (4)$$

As explained in the general formation of panel data, the equation demonstrates the dependent variables included in the  $Y_{it}$  model, the independent variables  $X_{it}$ , the initial slope parameter  $\beta_0$ , the known errors  $u_{it}$ , and the unit error  $\mu_{it}$ .

Numerous studies in the literature reveal that it is possible to calculate cross-sectional dependence using many methods. Breusch-Pagan (1980) Lagrange Multiplier (LM) test, Pesaran's (2004) LM test for Cross-Sectional Independence (CDLM) test, Pesaran's (2004) CD test and Pesaran, Ullah, and Yamagata's (2008) LMadj tests are often used in determining cross-sectional dependence. It is also notable that Friedman and Frees' test statistics are favored in determining cross-sectional dependence. Breusch and Pagan's (1980) LM test intends to test the correlation between units in the fixed effects model in the case of  $T > N$ , and it further tests whether there is a correlation between the cointegration or error correction models established for each unit (Yerdelen, 2017).

The stationarity assumption is applicable for models containing a time series and implies that the obtained mean, variance, and covariance values of the variables need to be independent of each other. The non-stationarity of the series reveals the existence of dependent variables, causing misleading results. Once the heterogeneity of data is ensured, the data can be considered suitable for panel data

analysis (Yerdelen, 2017). Since the panel regression model contains the time variable, it is necessary to perform a stationarity test, and if one finds an analysis that contradicts the stationarity assumption, the series can give insights into the period studied, but no interpretation can be made for other periods. The results of the analysis estimated by non-stationary series, relative to stationary series, cannot be explained by similar assumptions. The results obtained by the analysis applied to the model using non-stationary series may yield incorrect or inconsistent forecasting results (Aktaş, 2010).

The condition that a series is stationary can be indicated by the following equations:

$$E(Y_t) = \mu \quad (5)$$

$$VAR(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \quad (6)$$

$$E[(Y_t - \mu)(Y_{t-k} - \mu)] = \gamma k \quad (7)$$

Equations 5 and 6 indicate the variance and mean of the data in the time series; Equation 7 indicates that the series does not depend on a certain time (t) point, but on a time interval (k) between its two values (Baloch et al., 2019).

The term unit root, which is one of the terms of econometrics, means that the obtained time series data are not stationary; if a study detects a unit root, this implies that the series is not stationary. The Dickey-Fuller Test, which is about linear regression, can test the unit root. When a problem is detected in the correlation, the Augmented Dickey-Fuller Test is applied to the unit root tests. In addition to these tests, the Elliot-Lothman-Stock, Schmidt-Phillips, Phillips-Perron (PP), and Zivot-Andrews unit root tests are widely available in literature.

Another of the assumptions, cross-sectional data, is often used in social sciences and refers to the different unit values of the variables included in the model. It examines the changes in the relationships between the variables at the same time (Stock and Watson, 2011). This study draws on cross-sectional data by examining the population, GDP, and foreign trade deficit data for a certain period.

Panel data is defined as a data set in which the time series and the cross-section are integrated. While the cross-section provides information about different variables for the same period, panel data analysis also considers time. Thus, it contains information from multiple periods. The key difference that distinguishes panel data from cross-sections is that the same cross-section data is analyzed and interpreted at different periods (Wooldridge, 2019). A panel data analysis, using both time series and cross-section, yields the relationship and complexity between observations more reliably, it appears as a method commonly used in social sciences.

## 3.2. Forecasting Methods

### 3.2.1. The ARIMA Model

Developed the ARIMA method contains the autoregressive model (AR), the moving average model (MA), and the moving average model (SARIMA), which incorporates the periodic autoregressive model. The Augmented Dickey-Fuller unit root test intends to test whether the time series is stationary or not. In this method, differences between logarithmic transformations and references allow the stabilization of the data of the time series (Benvenuto, 2020).

Time series forecasting methods are mostly used for demand forecasting. The classical ARIMA approach becomes inapplicable when the stationarity assumption cannot be achieved after periodic corrections. For this reason, stationary variables are considered as the constraint of the ARIMA model. A large number of data are required for the applicability of the ARIMA method (Fattah et al., 2018).

An ARIMA model contains the values  $p$ ,  $q$ , and  $d$ . While the value  $p$  shows the number of autoregressive terms, the value  $d$  expresses the number of differences, and the value  $q$  indicates the number of moving averages.

As shown in Equation 8, the autoregressive process assumes that all these values have a linear function.

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t \quad (8)$$

The value  $\alpha_1$  in the equation is the regression coefficient. The behavior of time series can be influenced by the total effect of some processes. Although the GDP, population, and foreign trade deficit considered in this study are constantly changing, the effect on the total cargo forecast depends on the cumulative effect of the changes. Though the changes in the variables show an effect in the short term, their effect within the time series will be relatively low. That said, stationarity gains significance in performing statistical analysis of the time series. The method with the assumption that the change in the variables in two different periods is stationary, and Equation 9 demonstrates this integrated period.

$$Y_t = Y_{t-1} + \varepsilon_t \quad (9)$$

$\varepsilon_t$  in the equation is a normally distributed autocorrelation-free value with a mean of zero.

The moving average process is a linear combination of variation between values with one or more prior distortions. The order of the moving average contains the number of previous periods that are at the current value and is shown in Equation 10 (Liu et al., 2008).

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_l \varepsilon_{t-l} \quad (10)$$

The first step of the ARIMA method includes producing a stationary time series. This method can be applied with a time series with a stationary mean and autocorrelation, differentiation, or logarithmic transformation is required to balance the mean and variance of the data before taking any action on the model.

Once such transformation is performed, the next step is to determine the model. The parameters contained in the model are estimated in such a way as to reduce the overall error. In the last step, the adequacy of the model is assessed by testing the accuracy of the established hypotheses. Statistics and graphics for the results obtained reveal the future values. Different forecasting methods may be used where the adequacy of the model cannot be proven.

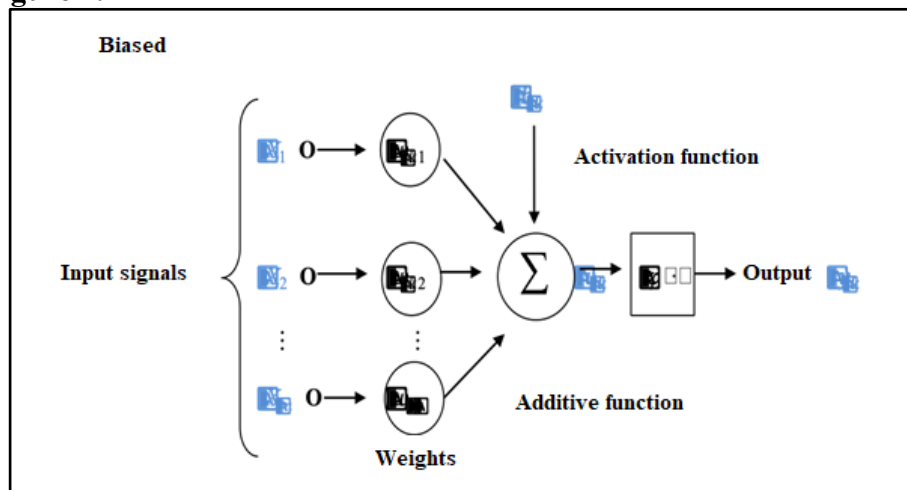
### 3.2.2. The Exponential Smoothing Method (ETS)

The ETS method is a method in which forecasts are constantly updated considering the change of data over the years. This method calculates the decreasing weights of the data of the time series for the most recent and oldest observations. In the data that are evaluated, the most recent data are considered as more weighted and relevant. This method, which can be applied to all series with both deterministic and stochastic trends, is usually used to make short-term forecasts (Jain and Mallick, 2017).

### 3.2.3. The ELM Neural Networks

Artificial neural networks are a prediction method developed by considering the structure and functioning of the human brain (Figure 1). The artificial neural networks method establishes a similar relationship to the relationship of the neurons in the nervous system connected by dendrites and axons. In this method, the addition function performs the functions of the dendrites; the activation function undertakes the task of the cells, and the weights carry out the task of synapses, that is, connection points.

**Figure 1.** The Structure of Artificial Neural Networks



Source: (Şen, 2004)

In the artificial neural network algorithm, neurons are monitored in three different layers. These are the input layer, the hidden layer, and the output layer. Single-layer artificial neural networks are not favored due to their shortcomings.

$$\left\| H \left( w_1, \dots, w_N, b_1, \dots, b_N \right) \hat{\beta} - T \right\| = \min_{\hat{\beta}} \left\| H \left( w_1, \dots, w_N, b_1, \dots, b_N \right) \beta - T \right\| \quad (11)$$

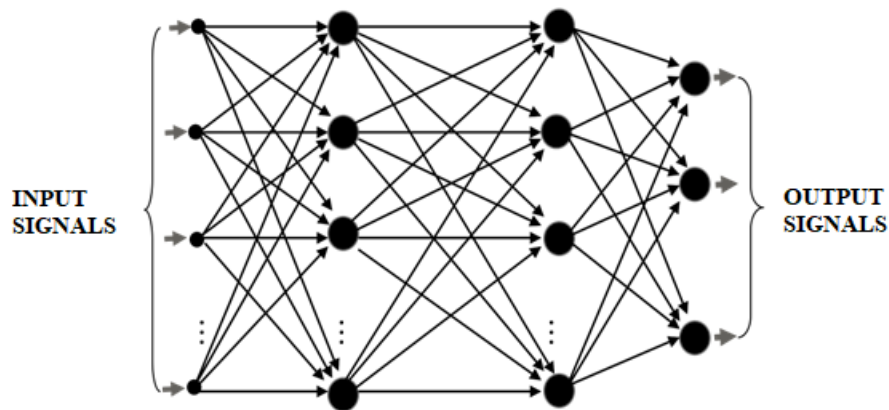
The ELM neural networks, which is the abbreviation of Extreme Learning Machine, denote fast learning machines. In the ELM, the weights of the inputs are indicated by  $w$ , and there is no need to adjust the deviations of the hidden layer.

According to the algorithm developed to train the fast-learning machine, finding  $\hat{\beta}$ , which is the least squares solution of the linear system  $H\hat{\beta} = T$ , is considered the necessary value for the training of the machine (Chen et al., 2019). The key advantage of this technique is that the least squares solution has a low margin of error. The random selection of weights and deviations in the ELM algorithm may lead to insolubility of the linear system, which can reduce the degree of accuracy of the forecast.

### 3.2.4. The MLP Neural Networks

In the MLP neural networks, called multilayer artificial neural networks, neurons are organized in the form of layers (Figure 2). Among these layers, the input layer ensures that the information needed for the solution is included in artificial neural networks. The layer where the data included in the system is processed and transmitted is the output layer. The layer between these two layers is called the hidden layer, and the MLP neural networks can have more than one layer.

**Figure 2.** The Multilayer Artificial Neural Networks



**Source:** (Şahin, 2018)

The goal of artificial neural networks is to learn the structure of variables by analyzing inputs and generalize for future impact. To achieve this, an artificial neural network monitors behavior and achieves the ability to generalize. For this algorithm to be used, the input and output values calculated earlier are defined depending on the time series of the system. In this method, which benefits from the back-propagation algorithm due to its comprehensibility and provability compared

to other methods, the output values given are called supervisors. Weights at the learning step are regulated by minimizing the error function using the formula in Equation 12 (Kaynar and Taştan, 2019).

$$E = \frac{1}{2} \sum_{k=1}^m (y_k - t_k)^2 \quad (12)$$

The MLP networks generally aim to change the network weights by applying backward propagation by minimizing the error between the intended output and the produced output until the most accurate result is found. Equation 13 shows the relationship between the output value obtained and the inputs from previous periods.

$$\gamma_t = w_0 + \sum_{j=1}^p w_j f \left( w_{oj} \sum_{i=1}^N v_{ij} y_{t-i} \right) + e_t \quad (13)$$

The values  $w_j$  and  $v_{ij}$  in the equation indicate the weight between nerve cells;  $p$  indicates the number of hidden nerve cells; and  $f$  refers to the activity function used in the hidden layer (Çuhadar, 2013).

### 3.2.5. TBATS

In the TBATS model, the letter T represents trigonometric terms for periodicity; the letter B represents Box-Cox transformations for heterogeneity, and the letter A represents ARIMA. The letter T in the rest of the model refers to trends, and the last letter, the letter S, refers to seasonal periods. Originally proposed by De Livera (De Livera and Hyndman, 2011), it can be expressed by the following equations:

Box-Cox transformation refers to the process of correcting irregular skewed data. It includes the identification of data that do not show a normal distribution by the lambda parameter and Equation 14 shows its formula.

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega} - 1}{\omega} & \text{if } \omega \neq 0; \\ \log y_t & \text{if } \omega = 0. \end{cases} \quad (14)$$

Seasonal periods are as follows:

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^M S_{t-m_i}^{(i)} + d_t \quad (15)$$

Global and local trends are:

$$l_t = l_{t-1} + \phi b_{t-1} + a d_t \quad (16)$$

$$b_t = (1 - \phi) b + \phi b_{t-1} + \beta d_t \quad (17)$$

ARIMA error is:

$$d_t = \sum_{i=1}^p \phi_i d_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (18)$$

Seasonal Fourier terms are:

$$S_t^{(i)} = \sum_{j=1}^{k_i} S_{j,t}^{(i)} \quad (19)$$

$$S_{j,t}^{(i)} = S_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + S_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \quad (20)$$

$$S_{j,t}^{*(i)} = -S_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + S_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \quad (21)$$

$$\omega = 2\pi j/m_i \quad (22)$$

In the formulas above,  $\omega$  refers to the Box-Cox transformation;  $y_t^{\square}$  represents the observation at time t.  $l_t$  is the local level at period t;  $b_t$  refers to the short-term trend in period t. While  $S_t^{(i)}$  indicates the seasonal component at time t,  $d_t$  shows the ARMA(p,q) process. T and m are seasonal periods, whilst T refers to seasonal patterns. Further,  $\lambda$  shows the parameter value,  $m_i$  is the period length of period i and k is the number of harmonic periods of the period i. Lastly,  $\theta$  and ARIMA(p,q) are coefficients;  $\varepsilon_t$  refers to a normally distributed value with a mean of zero according to the Gauss method.

Equation 23 demonstrates what the TBATS model includes.

$$TBATS(\omega, \emptyset), p, q, \{m_1, k_1\}, \{m_2, k_3\}, \dots, \{m_T, k_T\} \quad (23)$$

### 3.3. Research Model

This research draws on two dependent (Total Cargo, Total Container) and three independent variables (GDP, Population, Foreign Trade Deficit=Import-Export). The analysis step considers the total cargo and total container values as dependent variables to create two different panel regression models. At this step, all the necessary assumptions (stationarity, cross-sectional dependence, etc.) are tested through the analysis. As this study finds that the literature reports import and export data are interdependent, the variables are not considered alone, but foreign trade deficit data are integrated in this study. In obtaining the relevant data, this study benefits from the resources of the TÜİK and the Republic of Türkiye, Ministry of Transport and Infrastructure and considers the data of the TÜRKLİM Sector Report. Drawing on the data of all the provinces in the Marmara Region and the port authorities affiliated to these provinces, it seeks to determine the effect of the variables on the total cargo and total container demand, based on the data of all dependent and independent variables from 2012 to 2021.

Following the panel regression models, proceeding from part to whole, this study obtains the total data of the Northeastern Marmara region and forecasts the total cargo and container values using time series prediction techniques. Among various prediction techniques, ARIMA, ETS, ELM neural networks, and MLP neural network methods are applied to make comparisons through different prediction methods without relying on a single prediction method. All analyses are carried out on R software through specific codes.

## 4. Findings

### 4.1. The Data Set

This study draws on the information provided by the TÜİK for data related to GDP, Import, Export, and Population information of the provinces of the Marmara Region, and the 2022 statistics of the Republic of Türkiye, Ministry of Transport and Infrastructure for data on the port authorities and cargo handling of the ports in this hinterland (Ministry of Transport and Infrastructure, 2023.). The data were grouped for each province, and the provinces with no coastline were excluded from the analysis. Also, the provinces that do not have any container ports, perform container handling via their RO-RO ports temporarily in the years included in the data set, and accommodate ports for temporary transportation, were considered as outliers, and thus excluded from this study not to obtain erroneous results. To prepare the data for panel data analysis, the data were listed by period and by province, and the data from 2012 to 2021 were included. Remarkably, this study labels two of the data (total cargo, total container) as dependent and three (GDP, Population, Foreign Trade Deficit=Import-Export) as independent variables and then performs the analysis.

### 4.2. Findings from Panel Data Analysis and Prediction Techniques

In this section, panel regression analyses and time series prediction techniques are applied using research variables. For the analysis, the total cargo, total container, gross domestic product, population, and the export-import coverage ratio production for the years between 2012-2021 (T=10) in Balıkesir, Bursa, Çanakkale, İstanbul, Kocaeli, and Tekirdağ provinces (N=6) were used as the research variables. As the value ranges of the data in this study are high, analyses were made on the variables through the total cargo (LTCARGO), total container (LTOPCONT), gross domestic product (LGDP), and population variables (LPOPUL), which were converted into natural logarithm form.

This study created two different panel regression models, drawing on the data of gross domestic product, population, and export-import coverage ratio as independent variables, and the total cargo and total container data as dependent variables. As the data set used for the panel data analysis complies with the micro panel structure, the stationarity of the research variables was not examined (Baltagi, 2015; Yıldız and Demireli, 2019).

The findings of the analysis first presented the descriptive statistics for the research variables. In the next step, this study examined cross-sectional dependence. As  $T > N$  was found for the number of province and year variables included in the data while examining the cross-sectional dependence, the Lagrange-Multiplier (LM) test was applied (De Hoyos and Sarafidis, 2006). Dependencies were assessed by both the model and the variable.

While estimating the panel regression models through the total cargo and total container variables, a selection was performed between the pooled least squares (LS), random effect, and fixed effect type. Such selection was made using the F-test, Breusch-Pagan LM test, and Hausman test (Hausman, 1978; Hausman and Taylor, 1981). Following the identification of the most appropriate effect types, both models were checked against the changing variance (heteroscedasticity) and sequential dependence (autocorrelation) problems. The Breusch-Pagan (BP) test was applied to examine the changing variance problems in the models, and the Breusch-Godfrey (BG) test was applied to examine the sequential dependence problems (Torres-Reyna, 2010).

When any cross-sectional dependence, autocorrelation, or heteroscedasticity problem was detected in panel regression models established through the total cargo and total container variables, the robust Driscoll-Kraay (Driscoll & Kraay, 1998) covariance estimator was used for the significance of the model coefficients (Hoechle, 2007).

Following the panel data analyses, time series prediction models were used to predict the future data of the total cargo and total container volumes, which were evaluated separately as raw for the six provinces. For prediction models, ARIMA, exponential smoothing (ETS), TBATS, multilayer neural networks (MLP), and extreme machine learning (ELM) neural network techniques were applied. By that time, 70% and 30% of the total cargo and container data were allocated to the training and test, respectively, and the performance of the models was examined based on the error-squared averages of the test data. The performance of the models used to estimate the total cargo and container data on a provincial basis was evaluated based on the mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) criteria. The final forecasts of the total cargo and total container data for the coming years were obtained through the prediction models with the most successful performance.

**Table 1:** Descriptive Statistics of the Research Data.

Variable	Mean	SD	Min	Max	Skew	Kurt
LTCARGO	6.141	1.374	3.413	7.570	-0.720	1.964
LTOTCONT	5.113	1.376	2.255	6.553	-0.686	1.910
LGDP	7.665	0.544	6.912	8.787	0.824	2.745
LPOPUL	6.282	0.464	5.693	7.200	0.855	2.750
EICR	1.093	0.419	0.370	2.508	0.859	3.899

Mean: Mean, SD: Standard deviation, Min: Minimum, Max: Maximum, Skew: Skewness, Kurt: Kurtosis

**Source:** Authors

In this research, the margin of error for the test results was used as 1%, 5%, and 10%. The findings from the econometric analysis were obtained using the plm (Croissant and Millo, 2018), lmtest (Zeileis and Hothorn, 2002), forecast (Hyndman and Khandakar, 2008) and nnfor (Kourentzes, 2022) packages in the R program.

Table 1 shows the descriptive statistics of the six provinces for the total cargo, total container, gross domestic product, population, and export-import coverage ratio variables between 2012 and 2021. The mean LTCARGO is 6.141; the mean LTOPCONT is 5.113; the mean LGDP is 7.665; the mean LPOPUL is 6.282, and the mean EICR is 1.093.

**Table 2:** Results of the Cross-Sectional Dependence Test.

	CD test
Model-1	108.820*** (0.000)
Model-2	107.090*** (0.000)
LTCARGO	13.730*** (0.000)
LTOTCONT	14.225*** (0.000)
LGDP	16.424*** (0.000)
LPOPUL	16.421*** (0.000)
EICR	14.094*** (0.000)

\*p<0.10, \*\*p<0.05, \*\*\*p<0.01

Source: Authors

Table 2 shows the results of the cross-sectional dependency tests obtained for the first model (Model-1), where the total cargo, total container, gross domestic product, population, and export-import variables between 2012 and 2021 for the six provinces are dependent variables, and the second model (Model-2), where the total container variable is dependent. The results imply that the cross-sectional dependence is statistically significant in panel regression models where the total cargo and total container variables are used as dependents (p<0.01). The cross-sectional dependence was found to be significant also for the total cargo, total container, gross domestic product, population, and the ratio of exports to imports variables in this study (p<0.01). These results indicate that there is a cross-sectional dependence both on the model and variables.

#### 4.2.1 Results of the Panel Regression Analysis where the Total Cargo Values are Dependent

**Table 3:** F-test results for Model-1.

Model	F-test
Model-1 (Total cargo dependent)	0.512 (0.859)

**Source:** Authors

Table 3 offers the results of the F-test applied to select the effect type between the pooled LS and the fixed effect as part of the panel regression model, in which the total cargo variable was used as the dependent. As the results of the F-test were not found statistically significant, the pooled LS effect type, instead of the fixed effect, was selected for use in the model where the total cargo variable was dependent ( $p > 0.10$ ).

**Table 4:** The results of the Breusch-Pagan LM test for Model-1.

Model	LM-test
Model-1 (Total cargo dependent)	2.860* (0.094)

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

**Source:** Authors

Table 4 shows the results of the LM-test applied to select the effect type between the pooled LS and the random effect as part of the panel regression model, in which the total cargo variable was used as dependent. As the results of the LM-test were found as statistically significant, the random effect, instead of the pooled LS, was selected for use in the model where the total cargo variable was dependent ( $p > 0.10$ ).

**Table 5:** Results of the Hausman test for Model-1.

Model	The Hausman test
Model-1 (Total cargo dependent)	3.510 (0.319)

**Source:** Authors

Table 5 offers the results of the Hausman test applied to select the effect type between the random effect and the fixed effect as part of the panel regression model, in which the total cargo variable was used as the dependent. As the results of the Hausman test were not found as statistically significant, the random effect type, instead of the fixed effect, was selected for use in the model where the total cargo variable was dependent ( $p > 0.10$ ).

Assessing the results of the F-test, Breusch-Pagan LM test, and Hausman test, which were applied for the panel regression analysis, where the total cargo variable was used as dependent, the random effect type was preferred.

**Table 6:** Autocorrelation test for Model-1.

Model	The Breusch-Godfrey correlation test
Model-1 (Total cargo dependent)	40.700*** (0.000)

\*p<0.10, \*\*p<0.05, \*\*\*p<0.01

**Source:** Authors

Table 6 demonstrates the results of the autocorrelation test for the panel regression model where the total cargo variable was used as a dependent. The significance of the test statistics implies that this model has a significant autocorrelation problem (p<0.01).

**Table 7:** The Heteroscedasticity test for Model-1.

Model	The Breusch-Pagan heteroscedasticity test
Model-1 (Total cargo dependent)	8.984*** (0.029)

\*p<0.10, \*\*p<0.05, \*\*\*p<0.01

**Source:** Authors

Table 7 demonstrates the results of the heteroscedasticity test for the panel regression model where the total cargo variable was used as a dependent. Considering the significance of the test statistics, this model has a significant heteroscedasticity problem (p<0.01).

**Table 8:** The coefficient statistics of the panel regression model for Model-1.

Variable	Coefficient
Constant	-7.652*** (0.000)
LGDP	4.248*** (0.001)
LPOPUL	-2.917** (0.046)
EICR	-0.402 (0.234)

R<sup>2</sup>=0.661

Adj R<sup>2</sup>=0.643

\*\*\*p<0.01, \*\*p<0.05, \*p<0.1, R<sup>2</sup>: Coefficient of determination, Adj-R<sup>2</sup>: Adjusted coefficient of determination

**Source:** Authors

Table 8 presents the results of the panel regression model obtained using the Driscoll-Kraay estimator, resistant to cross-sectional dependence and random effect, that was used to determine the factors influential on the total cargo variable. Considering the regression coefficients of this model, the gross domestic product and population variables have a statistically significant effect on the total cargo data in six provinces ( $p < 0.05$ ). On the other hand, in this model, the export-import coverage ratio does not have a statistically significant effect on the total cargo variable ( $p > 0.10$ ). Given the regression coefficients in this model, it is notable that a 1% increase in gross domestic product value will lead to a 4.248% increase in the total cargo value. Further, a 1% increase in the population leads to a decrease of 2.917% in the total cargo volume.

#### 4.2.2. The Results of the Panel Regression Analysis where the Total Container Values are Dependent

**Table 9:** F-test results for Model-2.

Model	F-test
Model-2 (Total container dependent)	0.453 (0.898)

**Source:** Authors

Table 9 shows the results of the F-test applied to select the effect type between the pooled LS and the fixed effect as part of the panel regression model, in which the total container variable was used as dependent. As the results of the F-test were not found as statistically significant, the pooled LS effect type, instead of the fixed effect, was selected for use in the model where the total container variable was dependent ( $p > 0.10$ ).

**Table 10:** Results of the Breusch-Pagan LM test for Model-2.

Model	LM-test
Model-2 (Total container dependent)	3.180* (0.074)

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

**Source:** Authors

Table 10 offers the results of the LM-test applied to select the effect type between the pooled LS and the random effect as part of the panel regression model, in which the total container variable was used as dependent. As the results of the LM-test were found as statistically significant, the random effect, instead of the pooled LS, was selected for use in the model where the total container variable was dependent ( $p > 0.10$ ).

Table 11 offers the results of the Hausman test applied to select the effect type between the random effect and the fixed effect as part of the panel regression model, in which the total container variable was used as dependent. As the results of the Hausman test were not found as statistically significant, the random effect

type, instead of the fixed effect, was selected for use in the model where the total container variable was dependent ( $p > 0.10$ ).

**Table 11:** Results of the Hausman test for Model-2.

Model	The Hausman test
Model-2 (Total container dependent)	3.128 (0.372)

**Source:** Authors

Assessing the results of the F-test, Breusch-Pagan LM test, and Hausman test, which were applied for the panel regression analysis, where the total container variable was used as dependent, the random effect type was preferred.

**Table 12:** Autocorrelation test for Model-2.

Model	The Breusch-Godfrey correlation test
Model-2 (Total container dependent)	8.984*** (0.029)

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

**Source:** Authors

Table 12 demonstrates the results of the autocorrelation test for the panel regression model where the total container variable was used as a dependent. The significance of the test statistics implies that this model has a significant autocorrelation problem ( $p < 0.01$ ).

**Table 13:** The heteroscedasticity test for Model 2.

Model	The Breusch-Pagan heteroscedasticity test
Model-2 (Total container dependent)	41.167*** (0.000)

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

**Source:** Authors

Table 13 demonstrates the results of the heteroscedasticity test for the panel regression model where the total container variable was used as a dependent. Considering the significance of the test statistics, this model has a significant heteroscedasticity problem ( $p < 0.01$ ).

Table 14 gives the results of the panel regression model obtained using the Driscoll-Kraay estimator, resistant to cross-sectional dependence and random effect, that was used to determine the factors influential on the total container variable. Considering the regression coefficients of this model, the gross domestic product variable has a statistically significant effect on the total container data in

six provinces ( $p < 0.05$ ). On the other hand, in this model, the population and the export-import coverage ratio do not have a statistically significant effect on the total container variable ( $p > 0.10$ ). Given the regression coefficients in this model, it is notable that a 1% increase in gross domestic product's value will lead to a 4.386% increase in the total container value.

**Table 14:** The coefficient statistics of the panel regression model for Model-2.

Variable	Coefficient
Constant	-8.742*** (0.026)
LGDP	4.386* (0.087)
LPOPUL	-3.069 (0.311)
EICR	-0.444 (0.353)
R <sup>2</sup> =0.684	
Adj R <sup>2</sup> =0.668	

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ , R<sup>2</sup>: Coefficient of determination, Adj-R<sup>2</sup>: Adjusted coefficient of determination

Source: Authors

#### 4.2.3. Results of analyses of time series prediction model

**Table 15:** Performance criteria for the total cargo forecast of the Northeastern Marmara Region.

Model	Performance criteria		
	MSE	MAE	MAPE
ARIMA	18751940000000	4150081	8.361
ETS	3641923000000	1863773	3.798
<b>TBATS</b>	<b>2450369000000</b>	<b>1505043</b>	<b>3.072</b>
MLP	55339980000000	7390052	14.998
ELM	12851250000000	3364889	6.768

Source: Authors

Table 16 offers the performance criterion results of the models for five different time series prediction techniques for the total cargo data of the provinces

in the Northeastern Marmara hinterland. According to the results of the performance criteria, the lowest values of the total cargo data in the provinces in the Northeastern Marmara hinterland for the MSE (**245036900000**), MAE (**1505043**), MAPE (**3.072**) results, were obtained in the TBATS model. This finding implies that the TBATS model appears to be the most successful model for estimating the total cargo data for the provinces in the Northeastern Marmara hinterland as it has the lowest number of errors in predicting the total cargo data.

**Table 16:** Performance Criteria for the Total Container Forecast of the Provinces in the Northeastern Marmara Hinterland.

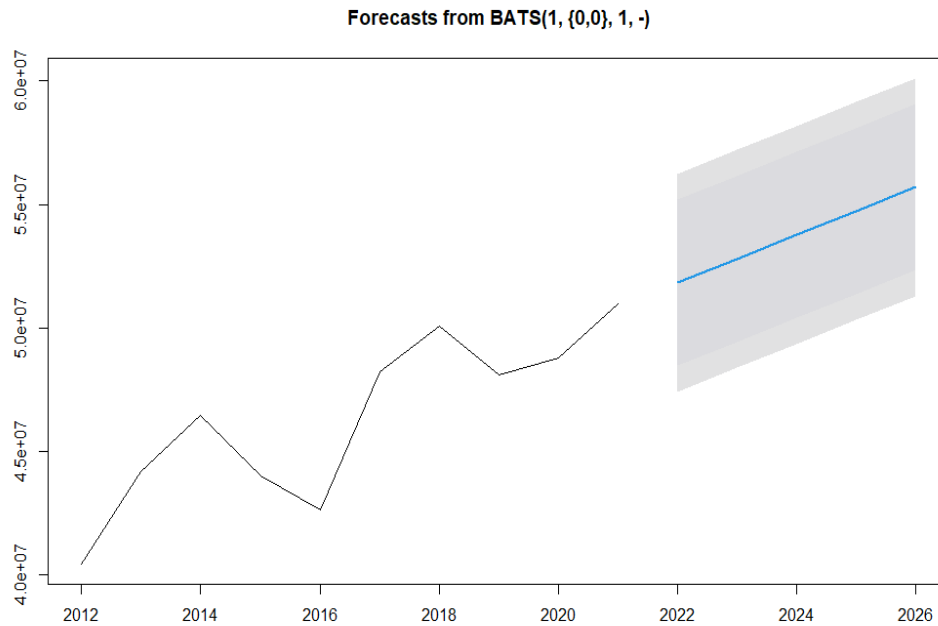
Model	Performance criteria		
	MSE	MAE	MAPE
ARIMA	301196155345	542080.000	11.164
ETS	10245417021	91379.650	1.900
<b>TBATS</b>	<b>10004742193</b>	<b>81972.100</b>	<b>1.703</b>
MLP	77742994244	256896.050	5.271
ELM	219540319809	460646.170	9.482

**Source:** Authors

Figure 3 shows the forecast results of the TBATS model for the next five years for the total cargo data of the provinces in the Northeastern Marmara hinterland. Accordingly, it is estimated that the total cargo volume in İstanbul and Kocaeli provinces will be on an upward trend in the next five-year period. Indeed, it is predicted that this trend will yield, respectively, 51829574, 52797178, 53764781, 54732385, and 55699989 tons for five years considering the total cargo amount in the Northeastern Marmara hinterland.

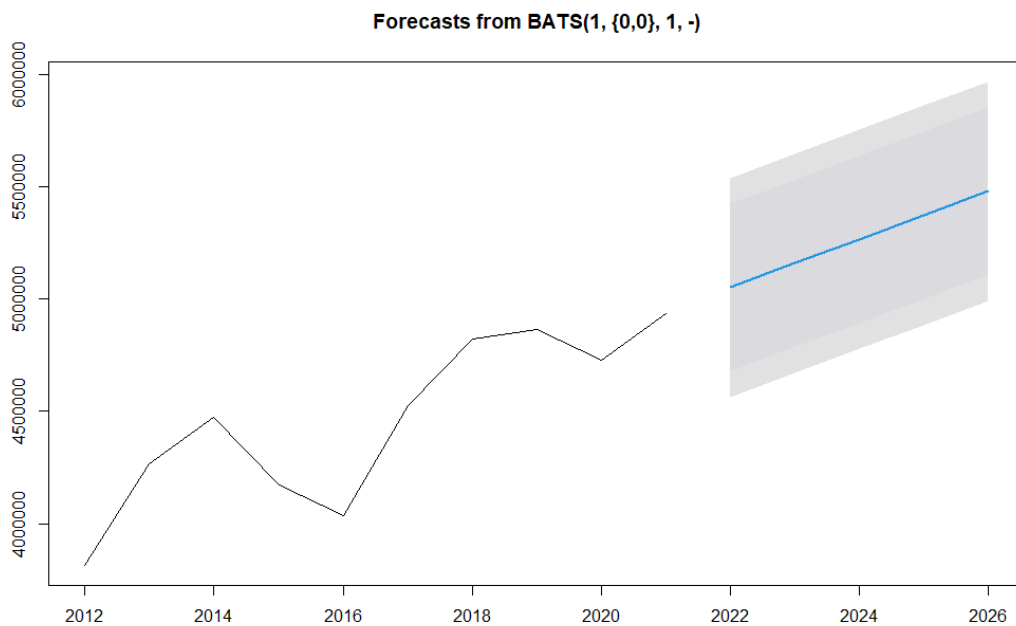
Figure 4 shows the forecast results of the TBATS model for the next five years for the total container data of the provinces in the Northeastern Marmara hinterland. Accordingly, it is estimated that the total cargo volume in the provinces in the Northeastern Marmara hinterland will be on an upward trend in the next five-year period. It is further foreseen that the total cargo volume of the provinces in the Northeastern Marmara hinterland will be, respectively, equal to 5052232, 5159448, 5266665, 5373881, 5481098 tons for the next five years.

**Figure 3:** Five-Year Total Cargo Forecasts of the Provinces in the Northeastern Marmara Hinterland.



Source: Authors

**Figure 4:** Five-Year Total Container Forecasts of the Provinces in the Northeastern Marmara Hinterland.



Source: Authors

## 5. Discussion, Conclusion and Suggestions

Using the dependent variables of total cargo and total container, this study assesses the effects of the variables of GDP, population, and the export-import

coverage ratio via the panel data regression analysis. GDP, which represents the market value of the final products produced by countries over a certain period, can be considered as the added value that countries offer to the market. Any increase in production leads to increased GDP, and the seaway is a frequently used means of transportation for the distribution of manufactured goods. Maritime transportation with its potential to carry more cargo at one time compared to other types of transportation, is expected to increase directly proportional to GDP. Indeed, this study determines that the 1% change in GDP will bring about an increase of 4.25 percent in the total cargo and 4.38 percent in the total container volume. The population variable in this study has an effect inverse to GDP. Despite the prediction that population growth increases the workforce, the final goods arriving in the region are consumed in this region due to the population causing a decrease in the total cargo and container volumes. This study reveals that a 1 percent change in the population will lead to a decrease of 2.97 percent in the total cargo amount and 3.06 percent in the total container amount. A key difference of this study from previous studies is that it considers the volume of imports and exports together. According to a large body of research conducted by economists, import and export figures are variables with interdependent characteristics. For this reason, this study considers imports and exports as the export-import coverage ratio. A one percent change in this ratio leads to a decrease of 0.4 percent in the total cargo and container volume. The negative effect of population on container demand may be explained by the consumption-oriented structure of densely populated regions. As the population increases, a significant share of goods may be consumed locally rather than transported through ports, reducing the volume of containerized shipments. Similarly, the negative impact of the export–import coverage ratio can be attributed to the structural characteristics of trade composition. Increases in this ratio may reflect a higher share of goods that are not suitable for containerization, such as bulk cargo, or changes in trade balance that reduce container flows. As for the suitability of the products imported and exported in Türkiye for container transportation, the sectors and transported goods relevant in this study on container transportation have an impact on this figure.

The second part of this study, performing analyses using five different techniques, estimates the total cargo demand of the provinces of the Marmara Region and the Northeastern Marmara hinterland. Based on the data between 2012 and 2021, this study ascertains that the trend identified has decreased the total cargo and container demand for the province of İstanbul, while it has boosted the cargo and container handled at the ports of Kocaeli. Specifically for the Northeastern Marmara Hinterland, it is estimated that this hinterland will enjoy approximately 55 million 700 thousand tons of cargo handling and 5 million 480 thousand TEU container handling 5 years later. This figure is 13% more than the cargo handled in the Northeastern Marmara in 2021. As for the container volume, by the end of 5 years, 14% more handling than the total TEU handled today is expected. This study finds the demand growth forecast for the northeastern Marmara ports, which is 11% in the 2022 TÜRKLİM sector report performed using linear regression, as 13%. Doğusel (2021) predicted a lower demand increase in container demand at the ports of Kocaeli in the upcoming period than this study estimates; to be precise, the

researcher forecasted an average container demand of 2.5 million TEU in 2025. This study finds this demand as 2.7 million TEU.

Previous studies report an increase at different rates but definitively argue that the container demand will increase in the future in all conditions. Considering the impact of developing technology and digitalization on every sector, it seems that port businesses in the region need to improve to meet such an increase. The common finding of all studies that the container demand at ports will increase, suggests the possibility that container ports may have difficulty in meeting the container demand in the future. Thus, ports have to turn to capacity-enhancing activities for the future intense demand. From a practical perspective, these findings suggest that port authorities and operators in the Northeastern Marmara region should prioritize capacity planning and infrastructure investments to accommodate the anticipated growth in container demand. Strategic actions may include expanding terminal areas, increasing handling equipment capacity, and improving hinterland connections. In addition, adopting data-driven planning approaches and integrating advanced forecasting techniques into decision-making processes can help reduce uncertainty and enhance operational efficiency. These measures are essential to ensure that ports can effectively respond to future demand increases and maintain their competitive position in the region. Given the urbanization in the region, it will not be enough for the sector that new businesses to develop, but public support will be needed as well. Based on the total cargo forecast of this study, it is notable that the level of capacity to meet demand will decrease soon in the absence of a suitable port area, and of businesses for enterprises to flourish and improve their operational processes. To close the container deficits, it is necessary to eliminate the uncertainties in the sector and direct investments correctly.

This study, which benefits from panel data analysis, evaluates the effect of variables and then performs demand forecasting through five different techniques. The use of five different techniques will allow us to find the technique that offers the most accurate forecast by considering the volume of the total cargo and container demand in the future. The most successful model in forecasting changes is based on data in the analysis of this study; for this reason, this study draws on multiple demand forecasting methods to reach the most accurate result. Performing demand forecast for the next five-year period, this study shows the increase of dependent variables on a year-by-year basis and reveals the need for businesses to respond to the demand that will increase in the short and long term.

Considering not only the ports located on the border of a port or a province, offering a total forecast of the hinterland, and using 5 different methods for such forecast, this study differs from the studies in the literature. Also, other studies in the literature disregard the dependency between the export and import variables, which are included as independent variables in regression analyses in other studies but are dependent on econometric studies. This study eliminates the dependency problem by including the export-import coverage ratio and this aspect makes it distinctive from previous studies.

Given that each region interacts with its hinterland, this study further reveals the need to conduct similar studies in other hinterlands and guide the businesses located in the regions.

This study contains some limitations that have been identified in previous studies and guides future research. First, the data obtained on a provincial basis constitute a limitation in assessing the port businesses as well as the hinterland. Though this study benefits from time series in its analysis step, the use of only three variables in the forecasting step raises the possibility of overlooking other factors that may affect demand. Even if the impact of this restriction has been reduced by applying a learning model to the dataset through artificial neural networks to achieve better results, evaluating maritime transportation, which represents a global sector, through three variables appears to be a limitation. The presence of the provinces that handle containers only during certain periods of the 10-year time series, and the presence of the provinces with no seashore in the hinterland, thus, no container handling data, are also considered as limitations for this study. To overcome this limitation, the provinces have been considered separately and a forecast has been made by taking into account the hinterland.

The lockdown caused by the COVID-19 pandemic, and the decreased import and export volume because of freight prices, also led to a deviation in the forecast demand performed in this study. The demand forecast in this study assumes that the trend formed by the influence of years will exhibit the same behavior periodically and in the future. In a nutshell, this study posits that the behavior seen over the last decade will be persistent in the future and that handling will not be positively or negatively affected by any kind of agreement, any war, or pandemic, whatsoever. For this reason, another limitation is that based on data on a provincial basis in this study, it is not possible to predict unusual developments that may occur in the coming years. The use of a longer time series to overcome this limitation may lead to a decrease in the margin of error. As statistical science and analytical elements such as demand forecasting become more significant for businesses, both public and private sector-based reporting will be high-volume, and research to be conducted in the coming years will allow the use of data containing a longer time series.

Considering that the volume of trade will increase the need for logistics will become constantly greater depending on imports and exports, and the countries will not be able to produce all by themselves, the figures of maritime transportation will increase every passing year. The demand for maritime transportation, which is much favored due to the low cost and the excess of cargo transported at once, will still attract scholarly research henceforth. Future studies need to increase the number of variables included in their regression model. Another suggestion is the use of more estimation methods to identify the correct method for estimation; new independent variables will, hopefully, reduce the error and deviation values in the forecasts.

## References

- Akar, O., & Esmer, S. (2015). Cargo Demand Analysis of Container Terminals in Turkey. *Journal of ETA Maritime Science*, 3(2), 117–122. Retrieved from <https://doi.org/10.5505/jems.2015.54254>
- Aktaş, C. (2010). Türkiye’de Reel Döviz Kuru ile İhracat ve İthalat Arasındaki İlişkinin VAR Tekniğiyle Analizi. *ZKU Journal of Social Sciences*, 6(11), 123–140.
- Baltagi, B. H. (2015). *The Oxford handbook of panel data*. Oxford University Press.
- Baloch, M. A., Zhang, J., Iqbal, K., & Iqbal, Z. (2019). The Effect of Financial Development on Ecological Footprint in BRI Countries: Evidence From Panel Data Estimation. *Environmental Science and Pollution Research*, 26(6), 6199–6208. Retrieved from <https://doi.org/10.1007/s11356-018-3992-9>
- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA Model on the COVID-2019 Epidemic Dataset. *Data in Brief*, 29, 105340. Retrieved from <https://doi.org/10.1016/j.dib.2020.105340>
- Chen, M.-R., Zeng, G.-Q., Lu, K.-D., & Weng, J. (2019). A Two-Layer Nonlinear Combination Method for Short-Term Wind Speed Prediction Based on ELM, ENN, and LSTM. *IEEE Internet of Things Journal*, 6(4), 6997–7010. Retrieved from <https://doi.org/10.1109/JIOT.2019.2913176>
- Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, 27(2). Retrieved from <https://doi.org/10.18637/jss.v027.i02>
- Çuhadar, M. (2013). Türkiye’ye Yönelik Dış Turizm Talebinin MLP, RBF ve TDNN Yapay Sinir Ağı Mimarileri ile Modellenmesi ve Tahmini: Karşılaştırmalı Bir Analiz. *Yaşar Üniversitesi E-Dergisi*, 8(31), 5274-5295.
- De Hoyos, R. E., & Sarafidis, V. (2006). Testing for Cross-Sectional Dependence in Panel-Data Models. *The Stata Journal: Promoting Communications on Statistics and Stata*, 6(4), 482–496. Retrieved from <https://doi.org/10.1177/1536867X0600600403>
- De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106(496), 1513–1527. Retrieved from <https://doi.org/10.1198/jasa.2011.tm09771>
- Doğusel, V. (2021). Kocaeli Limanları Talep Tahmini. *Deniz Taşımacılığı ve Lojistiği Dergisi*, 2(2), 82–90. Retrieved from <https://doi.org/10.52602/mtl.947252>
- Driscoll, J. C., & Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80(4), 549–560. Retrieved from <https://doi.org/10.1162/003465398557825>
- Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of Demand Using ARIMA Model. *International Journal of Engineering Business Management*, 10, 184797901880867. Retrieved from <https://doi.org/10.1177/1847979018808673>
- Hausman, Jerry A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251. Retrieved from <https://doi.org/10.2307/1913827>

- Hausman, Jerry A., & Taylor, W. E. (1981). Panel Data and Unobservable Individual Effects. *Econometrica*, 49(6), 1377. Retrieved from <https://doi.org/10.2307/1911406>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3). Retrieved from <https://doi.org/10.18637/jss.v027.i03>
- Jain, G., & Mallick, B. (2017). A Study of Time Series Models ARIMA and ETS. *SSRN Electronic Journal*. Retrieved from <https://doi.org/10.2139/ssrn.2898968>
- Kaynar, O., & Taştan, S. (2015). Zaman Serisi Aanalizinde MLP Yapay Sinir Ağları ve ARIMA Modelinin Karşılaştırılması. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*(33), 161-172.
- Kılınç, G., Karaatlı, M., & Ömürbek, N. (2022). Türkiye Limanlarındaki Konteyner ve Yük Elleçleme Hacimlerinin YSA NARX Modeli ile Öngörülmesi. *Verimlilik Dergisi*, (2), 251–266. Retrieved from <https://doi.org/10.51551/verimlilik.896525>
- Kourentzes, N. (2022). nnfor: Time Series Forecasting with Neural Networks. R package version 0.9.8. Retrieved 30 July 2024, from <https://cran.r-project.org/web/packages/nnfor/index.html>
- Liu, H., Tian, H., & Li, Y. (2012). Comparison of Two New ARIMA-ANN and ARIMA-Kalman Hybrid Methods for Wind Speed Prediction. *Applied Energy*, 98, 415–424. Retrieved from <https://doi.org/10.1016/j.apenergy.2012.04.001>
- Ministry of Transport and Infrastructure, R. of T. (2023). Konteyner İstatistikleri. Retrieved 27 July 2024, from <https://denizcilikistatistikleri.uab.gov.tr/konteyner-istatistikleri/>
- Pesaran, M. H. (2004). General Diagnostic Tests for Cross Section Dependence in Panels. CESifo Working Paper Series No. 1229; IZA Discussion Paper No. 1240. Available at SSRN: <http://ssrn.com/abstract=572504>
- Şen, Z. (2004). *Yapay Sinir Ağları*. Su Vakfı.
- Stock, J. H., & Watson, M. W. (2012). Dynamic Factor Models. In *The Oxford Handbook of Economic Forecasting* (pp. 35–60). Oxford University Press. Retrieved from <https://doi.org/10.1093/oxfordhb/9780195398649.013.0003>
- Solmaz, M. S., & Saygılı, M. S. (2008). Konteynerin Tarihçesi ve Konteyner Gemilerinin Gelişimleri. *Konteyner Deniz ve Liman İşletmeciliği (3-18)*. Beta yayınları.
- TÜRKLİM. (2022). *Turkish Port 2022 Sector Report*.
- Ubaid, A., Hussain, F., & Saqib, M. (2021). Container Shipment Demand Forecasting in the Australian Shipping Industry: A Case Study of Asia–Oceania Trade Lane. *Journal of Marine Science and Engineering*, 9(9), 968. Retrieved from <https://doi.org/10.3390/jmse9090968>
- Wooldridge, J. M. (2019). Correlated Random Effects Models with Unbalanced Panels. *Journal of Econometrics*, 211(1), 137–150. Retrieved from <https://doi.org/10.1016/j.jeconom.2018.12.010>
- Yaffee, R. A. (2003). A Primer for Panel Data Analysis. *Connect: Information Technology at NYU*, 8(3), 1–11.

- Yerdelen T. F. (2017). *Panel Zaman Serileri Analizi Stata Uygulamalı* (Vol. 3). İstanbul: Beta Basım Yayım Dağıtım.
- Yıldız, B., & Demireli, E. (2019). Nakit Yönetimini Etkileyen Değişkenler Üzerine BIST Perakende Sektöründe Panel Veri Analizi. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12(2), 416–429. Retrieved from <https://doi.org/10.17218/hititsosbil.589396>
- Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships.